

Supporting Information

Bejjanki et al. 10.1073/pnas.1417056111

SI Note 1: Experimental Procedures

Experiment 1: Performance in External Noise in AVGPs.

Participants. All 10 male NVGPs and 10 male AVGPs who participated in this experiment had normal or corrected-to-normal vision as measured with the logarithmic visual acuity chart “2000” (Early Treatment for Diabetic Retinopathy Study). All participants provided informed written consent and were paid \$8 per hour.

Participants completed the same video game playing questionnaire, aimed at establishing their video gaming history in the 12 mo before testing, as in ref. 1. Participants who reported playing at least 5 h of action video games per week during the past year were classified as AVGPs. Qualified action video games included first-person and third-person shooter games such as *Unreal Tournament*, *Counter Strike*, *Halo*, *Call of Duty*, and so on. Participants who had little to no action video game experience (less than 1 h per week) during the past year were classified as NVGPs; note that most NVGPs did play other kinds of games, such as board games, puzzle games, card games, strategy games, or social games.

Stimuli and procedure. Stimuli were displayed on a ViewSonic Graphics Series G225f CRT monitor with 1,024 × 768 pixel resolution and 120-Hz refresh rate, driven by a Mac OS X 10.4.11 system. A video switcher was used to combine two 8-bit output channels of the graphics card so that the display system could produce gray levels with 14 bits of resolution (2). Participants were tested in a dark room, with the mean display luminance set to 58 cd/m², and viewing distance set to 59 cm. Monitor gamma was calibrated by fitting the best power function to the measured luminance level (Minolta Chromameter, CS-100) of 10 different gray-level settings (from 0 to 240) of the monitor (full field).

The stimuli and task were similar to those of Lu and Doshier (3), except that the size (1.5 × 1.5°), degree of tilt (±2°), and reference angle (horizontal) were different. The stimuli were Gabor signals. The luminance distribution of the Gabor signal is described by the following equation:

$$L(x, y) = l_0 * (1.0 + c \sin(2\pi f(x \cos(\theta) + y \sin(\theta)))) * \exp(-(x^2 + y^2)/2\sigma^2),$$

where $\theta = \text{rad}(0^\circ \pm 2)$, l_0 is the background luminance, $f = 2$ cycles/degree is the center spatial frequency of the Gabor signal, $\sigma = 0.375^\circ$ is the SD of the Gabor's circular Gaussian envelope, and peak contrast c was determined by the adaptive staircase procedures (described below). External noise images consisted of pixels with contrasts independently drawn from an identical Gaussian distribution. To increase the noise energy in the task-relevant spatial frequency channels, external noise images were filtered by a band-pass filter with spatial frequencies ranging from one octave below to one octave above the spatial frequency of the Gabor signals. The sizes of both the Gabor signals and the external noise images, in terms of visual angle, were set to 1.5 × 1.5°.

A typical trial is shown in Fig. 1A of the main text. After a central fixation cross was presented for 250 ms a Gabor signal frame was presented, sandwiched between two external noise frames, for 16.7 ms per frame. The Gabor signal was tilted 2° clockwise or counterclockwise from 0°, with participants indicating the orientation of the Gabor—clockwise or counterclockwise from horizontal—using a key press. Participants received auditory feedback on their performance. Eight external noise contrast levels (0, 2.1, 4.1, 8.3, 12.4, 16.5, 24.8, and 33%) were used in an interleaved fashion, and for each level of external noise two interleaved staircase procedures were used, with step sizes of 10% of the

current contrast level, so as to converge upon the 79.37% or 70.71% signal contrast threshold (a three-down-one-up or a two-down-one-up staircase, respectively). Signal contrast threshold was defined as the mean of all of the reversals, excluding the first four. For each three-down-one-up staircase there were 100 trials and for each two-down-one-up staircase there were 80 trials, leading to a total of 1,440 trials per session. To improve our estimates, each participant carried out two such sessions in a row for a total duration of 1.5 h.

Experiment 2: Performance in External Noise After Action Video Game Training.

Participants. Thirty-six NVGPs (selected using the same video game experience criterion as in experiment 1 above) were initially recruited to participate in this training experiment and were divided randomly into an action group and a control group. However, not all recruited participants completed all parts of the experiment. The final sample included 12 action group participants (9 males and 3 females, 20–28 y old, mean age 23.9 y) and 14 control group participants (5 males and 9 females, 20–28 y old, mean age 22.6 y). Excluded participants included five female participants in the action group and two male participants in the control group who dropped out early during the video game training phase (described below) because of motion sickness or an inability to comply with the demanding schedule of the training, one male participant who was excluded from the control group owing to an inability to perform the noise-exclusion task at pretest, and two female participants who were excluded from the action group because one never learned to play the second action game used in the training study and the other did not conform to the prescribed schedule of training.

Video game training procedure. For both groups training consisted of playing two games for a total of 50 h (25 h per game). During the first 25 h of training, participants in the action group played the game *Unreal Tournament 2004* (Epic Games) in “Death Match” mode, where the goal is for the player to kill as many of the computer-controlled characters as possible while minimizing the number of deaths for the player-controlled character. During the second half of training (next 25 h of training), the action game trainees were switched to a second game, *Call of Duty 2* (Activision). This game places the player in fictionalized World War II combat situations, with the primary goal again being to kill as many computer-controlled characters as possible while minimizing deaths for the player-controlled character. Subjects were retested on all levels of *Unreal Tournament 2004* at the end of training to assess improvement in skill.

Participants in the control group played *The Sims 2* (2004, Electronic Arts Inc.) for the first half of training and *Restaurant Empire* (2003, Enlight Software Ltd.) for the second half of training. *The Sims 2* is a simulation-style game, wherein the player takes complete control over the simulated daily life of a character, from everyday activities (eating, bathing, etc.) to going to work, managing relationships with other characters, getting married, having and raising children, and eventual old age. As characters are added to the household the player takes control of those characters as well. *Restaurant Empire* is also a simulation-style game, wherein the player takes complete control of the simulated operation of restaurants, from developing recipes and menus to decorating restaurants, handling customers, and joining cooking competitions.

The training games for both the action and the control groups covered the entire display screen (visual angle of ~15° high × 18°

wide). Each group was requested to log their performance regularly (control game: accumulated wealth; experimental game: ratio of kills to death). In addition, because the action games did not control player advances in a systematic fashion, action group participants were required to reach a ratio of two kills for one death before they could advance to a harder level. This ensured that players progressed through the game smoothly, thereby avoiding long periods of frustration because the game became too hard or boredom because the game became too easy. This was not an issue with the control game, which progressed automatically through increasingly advanced situations.

Stimuli and procedure for pre- and posttest tasks. Participants were tested a few days before and a few days after the training period (designated as pretest and posttest, respectively) on the same orientation identification under external noise task as that used in experiment 1.

In addition, all participants were asked to fill out the “flow” questionnaire as defined by Csikszentmihalyi (4) after every 10 h of video game training, as well as after the pre- and posttest. This allowed us to assess participants’ engagement with the training games at different times during training. We were especially interested in determining whether the two training games were equally engaging. The specific questionnaire we used included 36 questions, covering four domains (enjoyment, attention, reward, and confidence). Participants answered each question by selecting one of five answers that ranged from “strongly disagree” to “strongly agree,” scoring 1–5, respectively.

RAPM. All participants were administered the RAPM at pretest to ensure participants across groups were matched in fluid intelligence, because differences in such aspects of intelligence have previously been related to differences in performance on perceptual tasks (5, 6). Raven scores, at pretest, were comparable between the two groups [mean score for AVGPs = 30 and mean score for NVGPs = 28.6, $t(20) = 0.85$, $P = 0.41$, two-tailed], indicating matched fluid intelligence across the two groups. Note that the exact same RAPM test was also administered at posttest. Both groups improved by a comparable amount from pre- to posttest [from a mean of 30–32.8 for AVGPs and from a mean of 28.6–29.4 for NVGPs, $F(1,20) = 5.31$, $P = 0.03$, $\eta^2 = 0.21$; no group effect, no interaction]. This pattern likely reveals a change in participants’ familiarity with the specific test items, rather than a true change in fluid intelligence (7, 8).

Finally, various pilot data were also collected from different subsets of the participants used here. All pilot experiments were administered after participants had completed the reported orientation identification task. One set of participants was tested on two visual working memory tasks, a task-switching paradigm, a speeded response time task, and visual tests focusing on center-surround interactions. A different set of participants was tested on a go/no-go task, the Posner letter identity task, a numerosity estimation task, an exogenous attention task, and an inhibition of return paradigm. The aim of these pilot experiments was to refine paradigms testing these various constructs in subsequent studies and will not be further discussed here.

Evaluating Long-Term Retention of Action-Trained Improvements in Performance (Posttest2).

Participants. Out of the nine action group participants and seven control group participants who were brought back to the laboratory a few months after the end of their training some had been trained 3 mo before, whereas others had been trained more than 1 y before (action-trained: five 3 mo, four more than 1 y; control-trained: three 3 mo, four more than 1 y). This schedule is a consequence of in-laboratory training studies’ being extremely time-intensive for participants. Accordingly, training studies are run only during the summer, so as to not interfere with participants’ academic performance. This training study was run over two consecutive summers, yet posttest2 for all participants occurred

at one point in time. Hence, some participants were run 3 mo after their last session (i.e., those subjects that had done the study during the previous summer) and others more than a year (i.e., those subjects that had done the study during the summer before that). This is also a way for us to check the durability of the effects over a relatively long time period. Participants were asked about their video game habits in the intervening period between the training study and the current test. All of the participants reported going back to their pretest habits in terms of video game play.

Stimuli and procedure. All of the participants were again tested (designated as post2-test) in the same orientation identification under external noise task as that used in experiments 1 and 2.

Experiment 3: Dynamics of Perceptual Template Learning in AVGPs.

Participants. All 10 NVGPs and 10 AVGPs who participated in this experiment had normal or corrected-to-normal vision, provided informed written consent, and were paid for their participation. All participants were selected using the same video game experience criterion as in experiment 1 above.

Stimuli and procedure. Stimuli were displayed on a Mitsubishi color graphic monitor (Diamond Pro-2070SB) with 1,024 × 768 pixel resolution, 85-Hz refresh rate, 22 inches in size and at a viewing distance of 58 cm. Subjects’ head position was stabilized using a chin and forehead rest. The stimuli and task were identical to those used in Jeter et al. (9), with the exception that only the high noise contrast (33%) and low precision ($\pm 12^\circ$) stimuli were used. Signal frames included a Gabor signal (subtending $3 \times 3^\circ$ visual angle, orientation $\pm 12^\circ$ around a reference angle of -35° or 55° , spatial frequency = 2 cycles/degree and SD of the spatial Gaussian envelope $\sigma = 0.4^\circ$). External noise frames included a noise patch, identical in size to the Gabor signal, made up of individual 2×2 pixel noise elements, with the contrast levels for the noise elements being drawn from a Gaussian distribution centered at the background luminance value and with an SD of 0.33 of 100% contrast. The Gabor signal and external noise patches were presented in the visual periphery (eccentricity = 5.67°) at one of two locations (in the northeast or southwest quadrants for half the participants and in the northwest or southeast quadrants for the other half). The reference orientation and the locations in which the stimuli were presented were randomly assigned to each subject and counterbalanced and matched across groups.

A typical trial is shown in Fig. 5A of the main text. After a central fixation cross was presented for 750 ms (the onset of each trial was signaled by an auditory tone), a Gabor signal frame was presented sandwiched between two external noise frames (30 ms per frame), leading to temporal integration of the Gabor signal with the external noise patches. Participants were tasked with indicating the orientation of the Gabor—clockwise or counterclockwise around the reference angle—using a key press. Auditory feedback was provided after the participants’ choice. Each participant carried out a total of eight sessions, four sessions per day over 2 d, with 312 trials per session. In each session, four randomly interleaved staircases—two-up-one-down and three-up-one-down staircases for each of the two stimulus locations—were used. Signal contrast thresholds were estimated by averaging all of the reversals in each staircase, excluding the first three reversals. Overall contrast thresholds for each participant were then computed by averaging the thresholds across all four staircases, thereby converging to the 75% correct threshold. In the first session, staircases quickly hovered around a signal contrast of about 0.77 for NVGP and AVGPs (Fig. S3). In subsequent sessions, to achieve efficient estimation, initial signal contrast levels for the two 2/1 and the two 3/1 staircases were computed by averaging the estimates from the corresponding 2/1 and 3/1 staircases from the immediately preceding session.

SI Note 2: Overview of the PTM

The PTM. Limited by nonlinear signal transduction as well as by many sources of variability, such as intrinsic stimulus variability, receptor sampling error, variability in neural responses, and loss of information during neural transmission, human perception exhibits many inefficiencies. Since the 1950s, sensory psychologists have adopted the equivalent input noise method, used by electrical engineers (10) to measure the intrinsic noise of amplifiers, to characterize the limiting noise in perceptual processes (11–15). In a typical application, this method systematically adds increasing amounts of external noise to the signal stimulus and observes how threshold (i.e., signal stimulus energy required for an observer to maintain a predetermined performance level) depends on the amount of external noise (Fig. S1). This threshold vs. external noise contrast (TvC) function can then be used to specify the sources of perceptual limitations in an observer model, which maps external stimuli to internal perceptual representations and predicts human performance using a decision process such as signal detection theory.

A number of components, derived from both sensory psychology and physiology, have been used to construct observer models, including a perceptual template, a nonlinear transducer, additive noise, multiplicative noise, and a decision rule (13, 14, 16–18). For example, internal additive noise is associated with absolute thresholds in perceptual tasks and internal multiplicative noise is associated with Weber's law behavior of the perceptual system. In the PTM (Fig. S2), perceptual inefficiencies are attributed to three limitations: internal additive noise, internal multiplicative noise, and perceptual templates (imperfectly) tuned to the target stimuli. Lu and Dosher (19) carried out an extensive analysis of all of the major existing observer models, including a linear amplifier model (13), a multiplicative noise model (20), and a multiplicative noise plus uncertainty model (17). The classical observer models are reduced cases of the PTM. The PTM accommodates all of the known standard properties of data in the equivalent input noise literature, as well as providing the best qualitative and quantitative account of a range of particular data sets.

The central idea behind the PTM is that of a filter or template through which the signal is passed so as to optimize pattern recognition; the efficiency of that filter is under the control of five main components (Fig. S2): (i) a contrast gain to the signal β that is normalized relative to its gain to the external noise, (ii) a nonlinear transducer function that raises its input to the γ^{th} power, (iii) a Gaussian-distributed internal multiplicative noise term with mean 0 and SD proportional to $(N_m \times)$ the contrast energy in the input stimulus, (iv) a Gaussian-distributed additive internal noise term with mean 0 and "constant" SD N_a , and (v) a decision process. In the PTM, accuracy of perceptual task performance is indexed by d' (18, 19):

$$d' = \frac{(\beta c)^\gamma}{\sqrt{N_{\text{ext}}^{2\gamma} + N_m^2 [N_{\text{ext}}^{2\gamma} + (\beta c)^{2\gamma}] + N_a^2}} \quad [\text{S1}]$$

For a given performance level, d' , we can solve Eq. S1 to express threshold contrast c_τ as a function of external noise (N_{ext}) in log form:

$$\log(c_\tau) = \frac{1}{2\gamma} \log \left((1 + N_m^2) N_{\text{ext}}^{2\gamma} + N_a^2 \right) - \frac{1}{2\gamma} \log \left(1/d'^2 - N_m^2 \right) - \log(\beta). \quad [\text{S2}]$$

Mechanisms of Performance Improvement. In signal processing there are three ways to improve signal-to-noise ratio: amplification,

improved filtering, and modified gain control. Similar principles of contrast gain, retuning of cellular signal selectivity, and reduced contrast gain have also been demonstrated in single-unit neurophysiology (21). Motivated by the principles in signal processing and neurophysiology, Lu and Dosher (22) developed a theoretical framework based on the PTM to distinguish three mechanisms underlying performance improvements (22). Stimulus enhancement acts by multiplying the contrast of the input stimulus by a factor greater than 1, which is mathematically equivalent to internal additive noise reduction. The behavioral signature for this mechanism is performance improvement (reduced thresholds) in the region of low to zero external noise (Fig. S2B). Because it affects both the signal and the external noise in the input stimulus in the same way, this mechanism does not affect performance in high external noise conditions. External noise exclusion improves performance by focusing perceptual analysis on the appropriate temporal duration, spatial region, and/or content characteristics of the signal stimulus. This focus serves to eliminate/admit external noise from further processing. The behavioral signature for this mechanism is performance improvements in the region of high external noise (Fig. S2C). Internal multiplicative noise reduction reduces a noise source that is proportional to the energy in the input stimulus. The mechanism produces a behavioral signature of performance improvements in both high and low levels of external noise (Fig. S2D). Improved perceptual template, or what has been termed *reweighting*, results in a combination of reduced additive internal noise and improved external noise exclusion. This also has the behavioral signature of performance improvements in both high and low levels of external noise. It corresponds to the use of a more closely matched filter to the signal (Fig. S2). Finally, measuring TvC functions at two or more criterion performance levels makes it possible to resolve the individual contribution of each mechanism in situations where more than one mechanism may be operative (23).

In the PTM, performance improvements across groups, or time, are respectively modeled by multiplying N_a by a factor $A_a < 1.0$ for better stimulus enhancement, N_{ext} by a factor $A_f < 1.0$ for better external noise exclusion, and N_m by factor $A_m < 1.0$ for reduced internal multiplicative noise. If all three mechanisms are operative, the contrast threshold vs. external noise function for a PTM becomes

$$\log(c_\tau) = \frac{1}{2\gamma} \left\{ \log \left[(1 + A_m^2 N_m^2) A_f^{2\gamma} N_{\text{ext}}^{2\gamma} + A_a^2 N_a^2 \right] - \log \left[\frac{1}{d'^2} - A_m^2 N_m^2 \right] \right\} - \log(\beta), \quad [\text{S3}]$$

where c_τ denotes the predicted contrast threshold, N_{ext} is the SD of external noises, d' is the perceptual sensitivity of the observer, corresponding to the thresholds in the two-alternative forced-choice task. Note that both reduced internal multiplicative noise and better perceptual templates predict a downward shift of the whole function. However, as shown in Dosher and Lu (23), whereas a perceptual template change predicts a ratio of contrast threshold between the compared conditions to be similar across performance levels (79.37% and 70.71% accuracy), a reduction in multiplicative noise predicts a greater ratio at the more stringent performance level (79.37%).

The external noise approach has been used to characterize mechanisms of performance improvements and decrements in a wide range of applications, including attention (22, 24), perceptual learning (25–29), adaptation (30), amblyopia (31, 32), perceptual interaction (33), dyslexia (34), and visual memory (35). In many cases, a pure mechanism of stimulus enhancement/diminution or external noise exclusion/admittance has been identified, with mixtures of these two mechanisms also being observed in

some instances. The internal multiplicative noise reduction/increase mechanism has not been found empirically.

SI Note 3: Data Analysis

Experiment 1: Performance in External Noise in AVGPs. An omnibus $2 \times 2 \times 8 \times 2$ ANOVA was performed with action game experience (AVGP/NVGP) as a between-subject factor; run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors; and log signal contrast threshold as the dependent variable. As expected, significant effects of external noise level [$F(7,126) = 260.41, P < 0.001, \eta^2 = 0.94$] and performance level [$F(1,18) = 284.07, P < 0.001, \eta^2 = 0.94$] were observed, indicating lower signal contrast thresholds at low external noise levels and at the easier performance criterion level. No main effect of runs was observed [$F(1,18) = 1.82, P = 0.19, \eta^2 = 0.09$], nor did run interact with any other variable (all $P > 0.05$). Importantly, AVGPs showed overall lower contrast thresholds than NVGPs [$F(1,18) = 5.82, P < 0.05, \eta^2 = 0.24$], indicating better performance in that group (Fig. 1B in the main text). No other significant effect or interaction was observed (all $P > 0.05$).

Fitting the PTM. Signal contrast thresholds across the eight external noise levels and two performance levels, for both AVGPs and NVGPs, were fit with the PTM to derive TvC curves for both groups (see ref. 19 for a review describing the PTM model and refs. 22 and 25 for studies describing the use of PTM models in capturing the mechanisms underlying performance changes owing to training or attention). In the PTM, performance improvements can be attributed to one or more of three noise reduction mechanisms: improved external noise exclusion, improved internal additive noise reduction, and improved internal multiplicative noise reduction. As discussed in SI Note 2 above, the signature of an improved perceptual template is a combination of reduced internal additive noise and improved external noise exclusion, with the ratio of contrast thresholds between the conditions under consideration being similar across performance levels (that is, across the 79.37% and 70.71% accuracy levels in our study).

We compared the results obtained by fitting eight models, ranging from no change in any of the three noise-reduction mechanisms in AVGPs to the full model with changes in all three noise-reduction mechanisms in AVGPs. As discussed above, the full model then consists of four shared parameters (β , Nm, Na, and γ) across the two groups and three AVGP-specific noise parameters A_f , A_a , and A_m . To find the best-fitting model, we compare nested models in a model lattice. The comparisons are not strictly independent. The logic is the same as in stepwise regression, where comparing models with more and more regressors is not considered as involving independent comparisons. The model with the least number of parameters that was still statistically equivalent to the model in which all changes were allowed was considered to be the best-fitting model (see ref. 3 for a similar approach). Model fitting was carried out using a least-square method, and an F test over nested models was used to compare the different models. Formally, the F test can be described as

$$F(df_1, df_2) = \left[\frac{(r_{full}^2 - r_{reduced}^2)}{df_1} \right] / \left[\frac{(1 - r_{full}^2)}{df_2} \right],$$

where $df_1 = k_{full} - k_{reduced}$, and $df_2 = N - k_{full}$. N is the number of predicted data points (i.e., 32 eight external noise levels \times two levels of performance \times two groups) and k_s is the number of parameters in each model.

The best-fitting model was the model with a combination of increased additive internal noise reduction and improved external noise exclusion in AVGPs. This model has six parameters, four shared between AVGPs and NVGPs (β , Nm, Na, and γ values),

and two AVGP-specific noise values, A_f and A_a . With six parameters and $r^2 = 98.89\%$, this reduced model was statistically equivalent to the full model with all noise parameters changed [$F(1,25) = 0.00, P = 0.99$] and was significantly better than the model with no noise parameters changed [$F(2,26) = 69.92, P < 0.01$], the model with external noise exclusion only [$F(1,26) = 41.69, P < 0.001$], and the model with internal additive noise reduction only [$F(1,26) = 54.81, P < 0.001$]. The results of the model fitting are plotted in Fig. 1B of the main text (fit lines). In addition, the model with only a multiplicative noise change provided a significantly worse fit ($r^2 = 98.03\%$) than the full model [$F(2,25) = 9.68, P < 0.001$]. Because multiplicative noise reduction predicts a larger improvement at the 79.37% than at the 70.71% accuracy level, but an improved perceptual template does not, we computed the ratios of contrast thresholds at each accuracy level (23). Crucially, the ratios of contrast thresholds between AVGPs and NVGPs were not significantly different at the two performance levels: 1.26 ± 0.03 at 79.37% correct vs. 1.22 ± 0.02 at 70.71% correct. These results make multiplicative noise change an unlikely mechanism underlying the improvements in AVGPs (23) and instead support the improved perceptual template as the mechanism underlying the performance differences between the two groups.

Experiment 2: Performance in External Noise After Action Video Game Training.

Orientation identification task. A $2 \times 2 \times 8 \times 2$ ANOVA was performed with group (action/control) as a between-subject factor; test (pre/post), run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors; and log signal contrast threshold as the dependent variable. No main effect of run was observed [$F(1,24) = 1.41, P = 0.25, \eta^2 = 0.06$], nor did run interact with any other factor (all $P > 0.05$). Main effects of external noise level [$F(7,168) = 414.22, P < 0.001, \eta^2 = 0.95$], performance level [$F(1,24) = 297.34, P < 0.001, \eta^2 = 0.93$], and test [$F(1,24) = 15.65, P < 0.001, \eta^2 = 0.40$] were observed, indicating the expected effects of lower thresholds at low external noise levels and for the less demanding performance level, as well as at posttest. Crucially, a test (pre/post) \times group (action/control) interaction was found [$F(1,24) = 8.66, P < 0.01, \eta^2 = 0.27$], indicating larger improvement in contrast thresholds between pre- and posttest in action trainees than in control trainees. Finally, a weak group \times external noise level interaction [$F(7,168) = 2.35, P < 0.05, \eta^2 = 0.09$] indicated an overall advantage throughout the whole experiment in the action group at low external noise levels. No other effects were observed (all $P > 0.05$).

The significant test \times group interaction led us to separately analyze action and control trainees' performance in the task. For each group, a $2 \times 2 \times 8 \times 2$ ANOVA was performed with test (pre/post), run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors and log signal contrast threshold as the dependent variable. For control trainees there was no effect of run [$F(1,13) = 0.23, P = 0.64, \eta^2 = 0.02$], nor did this factor interact with any other (all $P > 0.1$). Main effects of external noise level [$F(7,91) = 184.50, P < 0.001, \eta^2 = 0.93$] and performance level [$F(1,13) = 149.90, P < 0.001, \eta^2 = 0.92$] were observed. No other interactions were significant (all $P > 0.3$). Importantly, there was no effect of test (see Fig. 2B in the main text), indicating no reliable change in performance from pre- to posttest [$F(1,13) = 0.57, P = 0.46, \eta^2 = 0.04$].

For the action group, there was no effect of run [$F(1,11) = 1.83, P = 0.20, \eta^2 = 0.14$], nor did this factor interact with any other factor (all $P > 0.1$). Main effects of external noise level [$F(7,77) = 237.11, P < 0.001, \eta^2 = 0.96$] and performance level [$F(1,11) = 165.61, P < 0.001, \eta^2 = 0.94$] were observed. Crucially, a main effect of test was observed (see Fig. 2A in the main text), reflecting a lower signal contrast threshold and therefore improved

performance at posttest relative to pretest [$F(1,11) = 21.46, P < 0.001, \eta^2 = 0.66$]. The effect of test marginally interacted with performance level [$F(1,11) = 5.12, P = 0.045, \eta^2 = 0.32$], indicating a different magnitude of improvement at the two levels of task performance with greater improvement at the hardest level. No other effect was observed (all $P > 0.1$).

Fitting the PTM. Signal contrast thresholds in both pretest and posttest, across the eight external noise levels and two performance levels, for the control and action groups were then fitted with the PTM. The lack of difference between control and action groups at pretest [$F(1,24) = 0.01, P = 0.93, \eta^2 = 0.00$] led us to constrain β , Nm, Na, and γ (a subset of PTM parameters) to be the same across the two groups. Control and action groups' data in both pre- and posttests were then fitted simultaneously forcing the same β , Nm, Na, and γ parameters but allowing changes in the three noise-reduction mechanisms separately for each group, to capture the impact of the type of game training. This generated 64 models that range from changes in all three noise-reduction mechanisms for both groups to no change in noise reduction in either group. The model with the smallest number of parameters changed that was still statistically equivalent to a model with all parameters changed was defined as the best model fit. The model that assumed increased internal additive noise reduction and improved external noise exclusion in the action group, and no change in noise reduction in the control group, after training, provided the best fit. With six parameters and $r^2 = 98.60\%$, this model is statistically equivalent to the most saturated model with all noise parameters changed [$F(4,54) = 1.50, P = 0.22$] and is significantly better than the model with no change in noise parameters [$F(2,58) = 108.54, P < 0.001$], the model with external noise exclusion only [$F(1,58) = 120.14, P < 0.001$], and the model with internal noise reduction only [$F(1,58) = 44.33, P < 0.001$]. In addition, the model with only multiplicative noise change provided a significantly worse fit ($r^2 = 97.90\%$) than the full model [$F(5,54) = 7.20, P < 0.001$]. Results from these model comparisons therefore support improved perceptual templates in action video game trainees, rather than an alternative explanation such as a change in multiplicative noise. Indeed, a change in multiplicative noise predicts greater pre- to posttraining contrast threshold ratio at the 79.37% performance level than at the 70.71% performance level. This was not reliably observed in our data (1.35 ± 0.04 at 79.37% correct vs. 1.25 ± 0.03 at 70.71% correct), thereby further reinforcing improved perceptual templates as the most likely mechanism underlying improved performance of action trainees. The results of the model fitting are plotted in Fig. 2 of the main text (fit lines).

Game training results. To quantitatively assess how participants improved their game play as a result of training, several measures were used.

For the action games, kills and deaths in each block were used to calculate a skill metric $S = ([\text{Kills} - \text{Deaths}]/[\text{Kills} + \text{Deaths}])$. The S score was measured at all six levels of difficulty at pretraining, after 25 gaming hours (after *Unreal Tournament* training) and posttraining (hours 49 and 50 of game training). As with the control game, we compared pre- and posttraining S scores. The S scores at all of the difficulty levels increased from 0.65 (level 1), 0.58 (level 2), 0.35 (level 3), -0.07 (level 4), -0.45 (level 5), and -0.61 (level 6) to 0.94 (level 1), 0.84 (level 2), 0.75 (level 3), 0.42 (level 4), 0.08 (level 5), and -0.11 (level 6). Increases were observed for all participants and at all levels of difficulty, indicating generally improved performance with training.

For the control games, money accumulated was a reliable measure because it increases with positive actions such as making good recipes or adding a member to the household and decreases with negative actions such as burning down one's restaurant or having a character die due to neglect. All participants showed an accelerating increase in accumulated wealth over the course of training, indicating an improvement in game performance with

training. The time course of the accumulation was well fit by a linear function: For *The Sims2*, wealth = $(4,919 * \text{training hour}) + 110, r^2 = 98.05\%$; for *Restaurant Empire*, wealth = $(30,757 * \text{training hour}) - 6,020, r^2 = 96.23\%$. Taken together, these results demonstrate that both groups showed improvement in their respective training tasks.

In addition, it is worth noting that the control group trainees managed an average of three different characters during their first 25 h of training and an average of four different restaurants during their second 25 h of training. Thus, the control-group trainees played/managed on average 3.5 times as many characters/restaurants, had more variable goals, and were exposed to more diversity in their environment than the action-group trainees. We chose to keep the action game trainees at a disadvantage for these factors to ensure that the amount of stimulating situations encountered could not easily explain any improvements in the action group, beyond what is seen with the control group.

Flow measure. To investigate whether action and control games were similarly engaging, flow scores at hours 20 (toward the end of game 1) and 50 (after game 2) were compared across action and control groups (Fig. S3). On average, participants in both groups showed a high degree of engagement (the higher the flow, the greater the engagement) with their assigned games. A 2*2 ANOVA was performed with group (action/control) as a between-subject factor, order (first/second) as a within-subject factor, and flow score as the dependent variable. Although the mean flow score was slightly larger for the control trainees than for the action trainees, we found no significant effect of group [$F(1,24) = 2.32, P = 0.16, \eta^2 = 0.08$] or order [$F(1,24) = 0.38, P = 0.54, \eta^2 = 0.02$], nor did we find a significant group \times test order interaction [$F(1,24) = 0.12, P = 0.73, \eta^2 = 0.01$]. Thus, better performance in the action-trained group cannot be attributed to greater engagement with their assigned game.

Evaluating Long-Term Retention of Action-Trained Improvements in Performance.

Orientation identification task. A 2*2*2*8*2 ANOVA was performed using the pretest and post2-test data for the subset of participants that were brought back to the laboratory, with group (action/control) as a between-subject factor; test (pre/post2), run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors; and log signal contrast threshold as the dependent variable. Main effects of test [$F(1,14) = 24.06, P < 0.001, \eta^2 = 0.63$], performance level [$F(1,14) = 257.47, P < 0.001, \eta^2 = 0.95$], and external noise level [$F(7,98) = 318.19, P < 0.001, \eta^2 = 0.96$] were observed and, more relevant to our aim, a marginally significant interaction between pre/post2 and group [$F(1,14) = 3.25, P = 0.09, \eta^2 = 0.19$] was noted. In addition, we also noted an interaction between performance level and external noise level [$F(7,98) = 2.47, P < 0.05, \eta^2 = 0.15$], indicating a greater difference in contrast thresholds between the two performance levels at low than at high external noise levels, and an interaction between external noise level and group [$F(7,98) = 3.51, P < 0.01, \eta^2 = 0.20$]. Across all test sessions the action group showed lower contrast thresholds at low external noise levels than the control group, but no difference between the two groups was noted at high external noise levels. Finally, unlike any of the previous results, run was seen to interact with several factors, leading to triple and quadruple interactions (all P at 0.05 level), which are difficult to interpret.

Because the pre/post2 \times group interaction was only marginally significant [$F(1,14) = 3.25, P = 0.09, \eta^2 = 0.19$], we confirmed the presence of a test effect within the action group. A 2*2*8*2 ANOVA using the pretest and post2-test data for the subset of action group participants that were brought back to the laboratory, with test (pre/post2), run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors and log signal contrast threshold as the

dependent variable, confirmed a main effect of test [$F(1,8) = 21.64, P < 0.01, \eta^2 = 0.73$; see Fig. 3 in the main text], indicating that the action video game-induced improvements in performance were retained several months after training. As expected, main effects of external noise level [$F(7,56) = 178.85, P < 0.001, \eta^2 = 0.96$] and performance level [$F(1,8) = 105.58, P < 0.001, \eta^2 = 0.93$] were also observed, indicating lower thresholds at low external noise levels and for the less-demanding performance level, as expected. No other main effect or interactions were significant (all $P > 0.3$). To better characterize performance improvements of action trainees across the different tests, we further examined how their performance changed between posttest (immediately after training) and post2-test (several months after the end of training) compared with between pre- and posttest (immediately before and after training). Because a smaller group of subjects participated in post2 compared with the initial pre/post tests (experiment 2, discussed above), $2 \times 2 \times 8 \times 2$ ANOVAs were again performed with tests (pre/post or post/post2) as a between-subject factor; run (first/second), external noise level (eight levels), and performance level (79.37% and 70.71%) as within-subject factors; and log signal contrast threshold as the dependent variable, for the subset of action trainees that were brought back to the laboratory. We confirmed that even with this restricted group of participants the main effect of test was significant between pre- and posttest [$F(1,8) = 20.47, P < 0.01, \eta^2 = 0.72$], but it was not between post- and post2-test [$F(1,8) = 0.24, P = 0.64, \eta^2 = 0.03$], thereby suggesting that the observed improvements of action returnees in post2 most likely occurred while they trained on the action game, and that these improvements were retained after the end of training, between post- and post 2-test.

For completeness, we report the $2 \times 2 \times 8 \times 2$ ANOVA, with pre- and post2-test data, performed on the subset of control-trained participants that were brought back to the laboratory. The main effect of test (pre/post2) was almost significant [$F(1,6) = 5.71, P = 0.054, \eta^2 = 0.49$; see Fig. S4], owing mainly to an improvement from pretest to post2 at low levels of external noise, an effect that was not noted between pre- and posttest in the larger sample of control-trained participants. Main effects of external noise level [$F(7,42) = 162.85, P < 0.001, \eta^2 = 0.96$] and performance level [$F(1,6) = 226.05, P < 0.001, \eta^2 = 0.97$] were also observed. Unlike any of the previous results, run was seen to interact with several factors (all $P = 0.05$), with the meaning of such interaction remaining difficult to interpret. As we did for action trainees, pre- and posttest changes were explored separately from post- and post2-test changes in the small group of control trainees that returned for post2-test. The same $2 \times 2 \times 8 \times 2$ ANOVAs as performed with the action group trainees above showed no significant test effect between pre- and posttest [$F(1,6) = 0.42, P = 0.54, \eta^2 = 0.07$] and showed a marginally significant effect between post- and post2-test [$F(1,6) = 3.89, P = 0.10, \eta^2 = 0.07$]. Thus, the improvements in low noise observed in control returnees at post2 most likely occurred between post and post2-test, with performance remaining stable between pre- and posttest. Why performance improved at low noise levels between posttest and post2-test remains unclear because all subjects reported having stopped playing video games after the end of their training.

Fitting the PTM. The data from the control and the action returnees at pretest and post2-test was fitted with the PTM model. Action and control returnees showed a significant noise \times group interaction [$F(7,98) = 2.40, P < 0.05, \eta^2 = 0.15$] in their pretest data, suggesting different patterns of pretest performance in the subset of participants that were brought back to the laboratory. This fact, combined with the very different pattern of changes across pre-, post-, and post2-test in each group of returnees led us to fit each group's data separately by keeping the β , Nm, Na, and γ parameters (a subset of PTM parameters; see ref. 3) for

each group constant from pre- to post2-test and considering the full range of models, from no change in noise reduction in either group in post2 to a change in all three noise parameters in post2. The procedure for model fitting and selection was identical to that described above. The model-fitting results for the action group are plotted in Fig. 3 of the main text (fit lines) and the results for the control group are plotted in Fig. S4 (fit lines).

For the action group, the model that assumes improvements in internal additive noise reduction and external noise exclusion at post2 provided the best fit. Specifically, in comparison with the pretest, the best-fitting model assumed a 17% improvement in external noise exclusion and a 28% increase in internal noise reduction in post2. With six parameters and $r^2 = 99.25\%$, this model is statistically equivalent to the most saturated model with all noise parameters changed [$F(1,25) = 0.34, P = 0.57$] and is significantly better than the model with no change in noise parameters [$F(2,26) = 97.07, P < 0.001$], the model with external noise exclusion only [$F(1,26) = 99.15, P < 0.001$], and the model with internal noise reduction only [$F(1,26) = 38.83, P < 0.001$]. The model with only a multiplicative noise change ($r^2 = 98.26\%$) provided a significantly worse fit than the most saturated model [$F(2,25) = 16.89, P < 0.001$]. These model comparison results help reject a multiplicative noise change explanation and are in accordance with the finding of comparable contrast ratio at 79.37% performance level (1.32 ± 0.04) and at 70.71% performance level (1.29 ± 0.04). Together, as in experiments 1 and 2, these results support improved perceptual templates as the most likely mechanism underlying the observed improvements in action trainees and show that this improvement is maintained for several months after the end of training.

For the control group, the model that assumes an improvement in internal additive noise reduction at post2 provided the best fit. Specifically, in comparison with the pretest, the best-fitting model assumed a 19% improvement in internal noise reduction in post2, with no change in external noise exclusion. With five parameters and $r^2 = 97.86\%$, this model is statistically equivalent to the most saturated model with all noise parameters changed [$F(2,25) = 0.36, P = 0.70$] and is significantly better than the model with no change in noise parameters [$F(1,27) = 24.60, P < 0.001$] and fits the data better than the model that assumes improvement in external noise exclusion alone ($r^2 = 96.38\% < r^2 = 97.86\%$ of best-fitting model). Moreover, the model with only a multiplicative noise change ($r^2 = 97.19\%$) provided a significantly worse fit than the most saturated model [$F(2,25) = 4.39, P < 0.05$], providing no support for the multiplicative noise model. Indeed, we did not observe a greater contrast ratio at the 79.37% performance level (1.14 ± 0.04) than at the 70.71% performance level (1.16 ± 0.04), as predicted by the multiplicative noise model. Improved internal noise reduction is the most likely mechanism underlying the improvements observed in the control group trainees at post2-test. It remains unclear why we observe internal noise reduction between the end of training and retest at post2 in controls trainees, whereas in action trainees we see performance changes only during the training period and these changes are best accounted for by a change in perceptual template.

Experiment 3: Dynamics of Perceptual Template Learning in AVGPs.

Orientation identification learning task. A 2×8 ANOVA was performed on the contrast thresholds across the eight sessions, with group (AVGPs/NVGPs) as a between-subject factor and session (1–8) as within-subject factor. Main effects of session [$F(7,126) = 13.42, P < 0.001, \eta^2 = 0.427$] and group [$F(1,18) = 7.54, P < 0.05, \eta^2 = 0.295$] indicated significant learning effects on this orientation identification task and generally lower contrast thresholds in AVGPs compared with NVGPs. Most importantly, a significant interaction between session and group indicated that the two groups differed across learning sessions [$F(7,126) = 3.16, P < 0.01, \eta^2 = 0.149$]. Post hoc tests confirmed comparable

performance across the two groups in the first session [$F(1,18) = 0.05$, $P = 0.825$, $\eta^2 = 0.003$; Fig. S5] but lower signal thresholds in AVGPs in subsequent learning sessions (all $P < 0.05$; all P corrected for multiple comparison with the Sidak correction).

Fitting the elaborated power function. Contrast thresholds across the eight learning sessions were fit with an elaborated power function $c(t)$ with three free parameters: (i) α , baseline contrast; (ii) λ , incremental threshold; and (iii) ρ , learning rate. Formally, the power function can be written as

$$c(t) = \lambda(t)^{-\rho} + \alpha,$$

where t is the learning session. For the two groups of subjects, the complete model has six parameters (one λ , ρ , and α per group). The model that postulates no change between the two groups has three parameters (λ , ρ , and α), because it assumes that λ , ρ , and α are the same across the two groups. Between the fully saturated model and the no-change model, a lattice of models with different numbers of parameters was explored. Similar to the PTM fitting procedure, goodness of fit was measured by the r^2 statistic, and a model lattice consisting of eight models, ranging from the no-change model to the fully saturated model, was used to select the best-fitting model based on the following F test:

$$F(df_1, df_2) = \left[\frac{(r_{full}^2 - r_{reduced}^2)}{df_1} \right] / \left[\frac{(1 - r_{full}^2)}{df_2} \right].$$

The model that had the smallest number of parameters and was statistically equivalent to the complete model was defined as the best-fitting model (9). The number of predicted data points in this case was 16 (2 groups \times 8 time points).

The model that assumed that only the learning rate ρ differed between the two groups provided the best fit. With four parameters and $r^2 = 97.45\%$, this model is statistically equivalent to the full model [$F(2,10) = 0.54$, $P = 0.60$] and is significantly better than the model with no change between groups [$F(1,12) = 297.27$, $P < 0.001$]. Furthermore, it fits the data better than the model that assumes only baseline contrast α changes ($r^2 = 86.97\% < r^2 = 97.45\%$ of best-fitting model), as well as the model that assumes only incremental threshold λ changes ($r^2 = 90.73\% < r^2 = 97.45\%$ of the best-fitting model).

SI Note 4: Neural Model

We adapted a recent probabilistic neural model of orientation selectivity (36) to our task and asked which type of network changes could best explain the performance improvements observed as a result of action video game training.

Stimulus Design. Network simulations were run using stimuli similar to those used in our behavioral experiments (described above). Each stimulus image included an oriented Gabor signal that was tilted $\pm 2^\circ$ around the horizontal reference, and was created by assigning grayscale values to image pixels, according to the following function:

$$Z(x,y) = z_0 * (1.0 + c \cos(2\pi f(x \cos(\theta) + y \sin(\theta)))) * \exp(-(x^2 + y^2)/2\sigma^2),$$

where $\theta = \text{rad}(0^\circ \pm 2)$, z_0 is the background grayscale value, $f = 2$ cycles/degree is the center spatial frequency of the Gabor signal, $\sigma = 0.375^\circ$ is the SD of the Gabor's circular Gaussian envelope, and c is the maximum contrast of the Gabor signal, as a proportion of the maximum achievable contrast. The Gabor signal extended over $2.3 \times 2.3^\circ$ of visual angle and was rendered on a 23×23 pixel grid. In addition, as in ref. 36, each stimulus image was padded with extra pixels (10 pixels at each end), set to the background value, to eliminate possible edge effects.

Pixel gray levels for the external noise were drawn from a Gaussian distribution with mean zero and SD depending on the experimental condition. As in our behavioral experiments, we used eight levels of external noise in which the SD of the distribution was varied as a proportion of the maximum achievable contrast. The effective noise contrast levels used were 0.005, 2.1, 4.1, 8.3, 12.4, 16.5, 24.8, and 33%. Each noise element included a single pixel and the noise gray level values were added to the stimulus gray level values on a pixel-by-pixel basis to generate the noise-injected image.

Network Model of Orientation Selectivity.

Network architecture. For a complete description of the network architecture and the response properties of units in each stage of the network, the reader is referred to Bejjanki et al. (36). There are two points to note regarding the current implementation. First, the behavioral task used here differed from the task simulated previously by Bejjanki et al. (36) in two major ways: (i) The Gabor patch appeared in central vision rather than in peripheral vision and (ii) the Gabor orientation and tilt to identify were different. As such, in adapting the model to the current task, some baseline network parameters were changed so as to match network performance to the pretest performance of our behavioral participants. These parameters were then fixed and the minimal number of parameter changes necessary to capture the action video game-induced improvements in behavior was explored. All relevant parameters and their values are described in the following sections. Second, a note of caution is necessary about the hypothesized neural locus of action game-induced improvements in human participants. Although we show that the behavioral improvements observed as a result of action video game experience can be captured in our neural implementation by changing the pattern of thalamo-cortical feed-forward connections, we do not intend to claim that this is the specific neural locus of action game-induced improvements in human participants. Rather, as described in the main text, our intent here is to show using a biologically realistic neural model of orientation selectivity that improving the quality of the inference during the processing of perceptual information results in improved behavioral performance, in line with the improvements observed as a result of action game training. Ultimately, the precise neural locus of action game-induced improvements in performance is likely to depend on the specific task used, the sensory modalities involved, and the nature of feedback and training received by gamers. Irrespective of the precise locus of the changes, however, we argue that playing action video games leads to improved performance by changing the pattern of feed-forward connections between neural layers so as to instantiate improved probabilistic inference.

Visual stage 1. As in Bejjanki et al. (36), the input layer consists of an array of ON center-surround cells and an array of OFF center-surround cells, each of which contains 529 units arranged in a 23×23 grid. The firing rate of a cell at location (x, y) is given by

$$r_{ON}(x,y) = G[r_{baseline} + r_{center}(x,y) - r_{surround}(x,y)]$$

$$r_{OFF}(x,y) = G[r_{baseline} - r_{center}(x,y) + r_{surround}(x,y)],$$

where with $\alpha = \{\text{center, surround}\}$,

$$r_\alpha(x,y) = q(c) K_\alpha \left(\int_{-\frac{w}{2}}^{\frac{w}{2}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{(x-x')^2}{2\sigma_\alpha^2}} dx' \right) \left(\int_{-\frac{l}{2}}^{\frac{l}{2}} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-\frac{(y-y')^2}{2\sigma_\alpha^2}} dy' \right).$$

The new parameters used were $\sigma_{center} = 0.12^\circ$, $\sigma_{surround} = 0.5^\circ$, $K_{center} = 18$, $K_{surround} = 21$, $r_{baseline} = 15$ Hz, and $q(c)$ is given by

$$q(c) = \left(\frac{\beta [\log(c)]^2}{c} \right) \text{ with } \beta = 0.1.$$

Visual stage 2. The stage 2 layer contains 1,024 neurons, which as in Bejjanki et al. (36) are modeled as LNP units. Specifically, each stage 2 cell is modeled as a single point process and the input–output relationship in the cell is composed of three distinct operations: (i) a linear step where the input spike trains, both feed-forward and recurrent, are linearly combined to obtain a “membrane potential proxy,” (ii) a nonlinear step where the membrane potential proxy is passed through a static nonlinearity to generate the instantaneous probability that it emits a spike, and (iii) a Poisson step where the instantaneous spike probability is used to generate Poisson spikes. These spikes are in turn transmitted to all of the other stage 2 cells through lateral connections, thereby influencing the postsynaptic cells’ membrane potential proxies at the next time step.

As in Bejjanki et al. (36), we model the receptive field of each cell in visual stage 2, with respect to visual stage 1, using a Gabor function $gab(x, y, \theta)$ defined as

$$gab(x, y, \theta) = e^{-\left[\frac{c_x^2}{2\sigma_x^2} + \frac{c_y^2}{2\sigma_y^2} \right]} \cos(2\pi k C_x),$$

where

$$\begin{aligned} C_x &= x \cos \theta + y \sin \theta \\ C_y &= y \cos \theta - x \sin \theta. \end{aligned}$$

The strength of each connection was set to $\alpha * |gab(x, y, \theta)|^2$. The new baseline parameters used were $\sigma_x = 0.25$, $\sigma_y = 0.3$, $k = 1.45$, and $\alpha = 0.55$.

As in Bejjanki et al. (36), we implement full lateral connectivity with the strength of the connection between two cells x and y , with preferred orientations P_x and P_y (in radians), given by

$$W(x, y) = \frac{G_w}{N_{out}} \left[e^{K_e(\cos(P_y - P_x) - 1)} - A_i e^{K_i(\cos(P_y - P_x) - 1)} \right] + DC_w.$$

The new baseline parameters used were $K_e = 0.5$, $K_i = 0.25$, $DC_w = -2.0$, $G_w = 1$, and $A_i = 0.4$.

Decision stage. The final stage of the network involves connections from all stage 2 cells to a single decision unit that outputs an estimate for the orientation of the stimulus (clockwise or counter clockwise around horizontal, for example). As in Bejjanki et al. (36), we used a linear classifier as the decoder in our network, which was optimized for the before-training network condition and then kept constant across all network manipulations.

Computing Orientation Discrimination Performance and Deriving TvC Curves. As in Bejjanki et al. (36), we compute the orientation discrimination performance of our network, when presented with the noisy oriented Gabor stimuli (described above), by estimating Fisher information, an information-theoretic quantity that directly predicts performance in discriminations tasks. Re-

cently, we have derived an analytic expression for the linear Fisher information in a population of LNP neurons driven to a noise-perturbed steady state, with network properties similar to those of the network used here (37). Linear Fisher information is the fraction of Fisher information that can be recovered by a locally optimal linear estimator—in practice, it has been found to provide a tight bound on total Fisher information, both in simulations (38) and in vivo (39). This expression, when applied to a network with a fixed decoder, can be written as

$$I(W_{dec}) = \frac{(W_{dec}^T \mu')^2}{W_{dec}^T \Gamma W_{dec}},$$

where

$$\begin{aligned} \mu' &= (D^{-1} - W)^{-1} M h' \\ \Gamma &= (D^{-1} - W)^{-1 T} [M \Gamma_{hh} M^T + D^{-1} G D^{-1}] (D^{-1} - W)^{-1}. \end{aligned}$$

W_{dec} represents the fixed pattern of connection weights from stage 2 to the decision stage, W represents the matrix of stage 2 recurrent connections, M represents the matrix of stage 1 – stage 2 feed-forward connections, h represents the mean input firing rates from stage 1, Γ_{hh} represents the covariance matrix of the input firing rates from stage 1, G is a diagonal matrix whose entries give the mean response of the LNP neurons, obtained from the nonlinear function that transforms the membrane potential proxy $u(t)$ to the firing rate for the LNP neurons, and D is a diagonal matrix that gives the derivative or slope of the activation function G of the stage 2 neurons at steady state.

Using the above expression allows us to compute the Fisher information, and hence the discrimination threshold, at the decision stage in response to a given stimulus. We compute Fisher information at the decision stage using stimuli with 17 signal contrast levels (10, 12, 14, 16, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, and 90%) and with the eight levels of external noise used in our behavioral experiments. We then derive network TvC curves by computing an iso-information contour for a value of information that is equivalent to the percent correct criterion used in the behavioral experiments (computed via signal detection theory), through the resulting information matrix (see ref. 13 for further details). We can now examine the influence of changes in network parameters on these network TvC curves in an effort to model the neural basis for the action-trained improvements in performance.

Modeling Action-Trained Improvements in Performance. As discussed in the main text (Fig. 4), action video game-induced changes in performance were captured in our model by making changes only to the feed-forward connectivity between stage 1 and stage 2. A change in this single network parameter led to a decrease in network signal contrast thresholds and a nearly uniform downward shift in network TvC curves (Fig. 4B), similar to those observed in our behavioral experiments. The specific network parameter values that led to the reported results are listed in Table S1.

- Hubert-Wallander B, Green CS, Sugarman M, Bavelier D (2011) Changes in search rate but not in the dynamics of exogenous attention in action videogame players. *Atten Percept Psychophys* 73(8):2399–2412.
- Li X, Lu Z-L, Xu P, Jin J, Zhou Y (2003) Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *J Neurosci Methods* 130(1):9–18.
- Lu ZL, Doshier BA (2004) Perceptual learning retunes the perceptual template in foveal orientation identification. *J Vis* 4(1):44–56.
- Csikszentmihalyi M (1990) *Flow: The Psychology of Optimal Performance* (Harper and Row, New York).
- Evans RB, Deary IJ (1994) Sensory discrimination and intelligence: Postmortem or resurrection? *Am J Psychol* 107(1):95–115.

- Melnick MD, Harrison BR, Park S, Bennetto L, Tadin D (2013) A strong interactive link between sensory discriminations and intelligence. *Curr Biol* 23(11):1013–1017.
- Detterman DK (1993) *Transfer on Trial: Intelligence, Cognition, and Instruction*, ed Sternberg RJ (Ablex Publishing, Westport, CT), p 296.
- Redick TS, et al. (2013) No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *J Exp Psychol Gen* 142(2):359–379.
- Jeter PE, Doshier BA, Petrov A, Lu Z-L (2009) Task precision at transfer determines specificity of perceptual learning. *J Vis* 9(3):1–13.
- North DO (1942) The absolute sensitivity of radio receivers. *RCA Review* 6(3):332–343.
- Barlow HB (1956) Retinal noise and absolute threshold. *J Opt Soc Am* 46(8):634–639.
- Nagaraja NS (1964) Effect of luminance noise on contrast thresholds. *J Opt Soc Am* 54(7):950–955.

